Trajectory similarity analysis in movement parameter space

Dodge, S ; Weibel, Robert ; Laube, P

Abstract: This paper introduces a similarity analysis method for moving object trajectories. The proposed method assesses the similarity between a set of trajectories in a multidimensional space, whose dimensions are formed by different movement parameters (e.g. position, speed, acceleration, direction), plus time. We investigate the applicability of the proposed method in finding relative movement patterns such as coincidence and concurrence in the movement of North Atlantic hurricanes.

Originally published at:
Dodge, S; Weibel, Robert; Laube, P (2011). Trajectory similarity analysis in movement parameter space. In: 19th GIS Research UK Annual Conference 2011, Portsmouth, 27 April 2011 - 29 April 2011. University of Portsmouth, 270-279.

Trajectory Similarity Analysis in Movement Parameter Space

Somayeh Dodge, Robert Weibel, Patrick Laube

University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
Tel. +41 44 63 55131, Fax +41 44 63 56848
{somayeh.dodge | robert.weibel | patrick.laube}@ geo.uzh.ch

ABSTRACT: This paper introduces a similarity analysis method for moving object trajectories. The proposed method assesses the similarity between a set of trajectories in a multidimensional space, whose dimensions are formed by different movement parameters (e.g. position, speed, acceleration, direction), plus time. We investigate the applicability of the proposed method in finding relative movement patterns such as *coincidence* and *concurrence* in the movement of North Atlantic hurricanes.

KEYWORDS: Trajectory similarity analysis, movement parameters, movement patterns, knowledge discovery, moving objects.

1 Introduction

In many domains of science and technology, understanding the collective movement behaviour of dynamic objects (i.e. humans, animals, vehicles, etc.) or processes (e.g. hurricanes) is very important. Nowadays, the advances in positioning technologies provide access to massive amounts of movement data in diverse application domains. The availability of such valuable repositories of movement data requires the development of new knowledge discovery tools in order to extract meaningful information and discover patterns of movement behaviours of mobile objects.

In order to study the dynamic behaviour of objects, it is necessary to observe the movement characteristics along the objects' geospatial lifelines, in addition to the positional information. These characteristics, so called *'movement parameters'* (MP) (Dodge et al., 2008), are divided into two types of 'instantaneous' parameters (i.e. detectable at individual moments) such as position, speed, and acceleration and 'relative' parameters (i.e. measurable over time intervals) such as relative speed, direction, and path sinuosity (Laube et al., 2007; Giannotti and Pedreschi, 2008).

However, in spite of the recent progress in the field of knowledge discovery and data mining (Giannotti and Pedreschi, 2008; Miller and Han, 2009), most of the existing spatio-temporal analysis techniques for moving object data deal only with the *positional* information of the tracked objects over time (i.e. with trajectory *geometry*), and very little attention has been paid to other movement characteristics. The same can be observed in the available literature on similarity analysis of movement data (Vlachos et al., 2002; Chen et al., 2005; Trajcevski et al., 2007; Pelekis et al., 2007; Buchin et al., 2009).

Similarity analysis is crucial in the process of knowledge discovery from movement data. The results of similarity analysis can significantly contribute to other important mobility data mining tasks, such as *trajectory classification* (Dodge et al., 2009), *trajectory clustering* (Zhang et al., 2006; Chen et al., 2005), or *movement pattern detection* (Gudmundsson and van Kreveld, 2006; Buchin et al., 2008).

The aim of this paper is to propose a spatio-temporal similarity analysis method with the perspective of detecting trajectories with similar dynamic behaviour. That is, the method assesses the similarity of the evolution of objects' movement parameters over time. The technique uses the Euclidean distance in a multidimensional space of movement parameters and can be applied for the detection of the movement patterns *coincidence* (i.e. similar positions over time) and *concurrence* (i.e. similar movement parameters over time). Such patterns occur when a set of objects exhibits a synchronous movement or at least similar movement parameters over a certain duration (Dodge et al., 2008). Similar to our approach, a number of previous studies used the Euclidean distance to assess the similarity of movements of objects (e.g. Yanagisawa et al., 2003; Buchin et al., 2009). However, these methods are based on the *positional* information of trajectories in the space-time cube. Therefore, these techniques can only detect *coincidence* patterns. To the best of our knowledge, no other authors so far have used the Euclidean distance in an n-dimensional movement parameter space. By doing so, however, our method is capable of detecting both *coincidence* and *concurrence* patterns.

2 Methodology

When an object moves about in space, the evolution of its movement parameters over time can be seen as functions over time, so-called *'movement parameter profiles'* (Dodge et al., 2009). The properties of these profiles can be quite different for different object types: Some may be rather smooth, others may express diversity in their evolution. However, when multiple moving objects form particular movement patterns such as *concurrence* or *coincidence*, their movement characteristics to a certain degree exhibit similar trends. Therefore, we may exploit information about the movement parameters of a given type of dynamic object for extracting spatio-temporal similarities among trajectories. Accordingly, in this paper, in comparison between two or more trajectories, *the movement characteristics of objects are considered similar when the evolution of their movement parameters resembles each other over a given period time.*

The methodology used in this study consists of four steps as presented in the following sections: 1) trajectory pre-processing, 2) set up a multidimensional movement parameter space, 3) similarity computation in the MP space, and 4) movement pattern detection.

2.1 Trajectory Pre-processing

Due to the nominal precision and accuracy of the positioning technologies used as well as the influence of the environment and other external factors, usually the raw movement data to some degree contain noise, gaps, and outliers. Therefore, to obtain more reliable trajectories for the purpose of similarity analysis, an initial stage of data cleaning and pre-processing (i.e. filtering, smoothing, resampling, etc.) is recommended, tailored to the peculiarities of the specific application domain. The aim is to eliminate the effects of noise, outliers, and other positioning errors from the raw movement data, and to generate regularly sampled trajectories of the same and synchronised duration.

2.2 Trajectory Representation in the Multidimensional MP Space

In the second stage, in order to assess the similarity between the movements of a set of objects, a required set of movement parameters for all fixes along the trajectories is computed. As an example, Figure 1.a illustrates the computed MP profiles from a sample set of speed (v), acceleration ($accl$), and direction (Az) for a trajectory. Then, a multidimensional MP space is generated from the computed movement parameters for each trajectory (Fig. 1.b). Each of

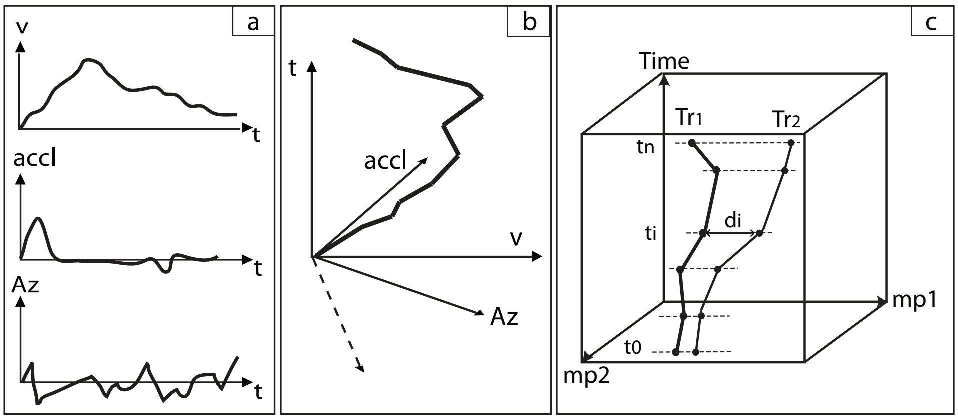


Figure 1: a) Computed MP profiles; b) the MP space; c) similarity computation in a 3D MP space

the movement parameters constitutes one dimension of the multidimensional feature space for trajectories. This provides a multidimensional profile (time series) for each trajectory.

The selection of movement parameters depends on the purpose of the similarity search study, and the type of movement characteristics that one wishes to compare among objects. For instance, speed (i.e. the rate of change of an object's position) and acceleration (i.e. the rate of change of an object's speed) give an indication of how slow or fast, smooth or jerky is the movement. Azimuth (i.e. direction of the movement) and turning angle (i.e. the change of direction) indicate the geometric shape of the trajectory, and the straightness index (i.e. the ratio of the length of the travelled path and the straight-line displacement) gives an indication of the sinuosity of the trajectory at a specific point.

2.3 Similarity Computation in the MP Space

In order to quantify the similarity between two trajectories of the same duration, the average Euclidean distance between the two multidimensional profiles is applied as the similarity measure. Equation 1 presents the computation of the distance between two trajectories Trj_1 and Trj_2, where $|Trj_1| = |Trj_2| = n$:

$$D(Trj_1, Trj_2) = \frac{\sum_{i=1}^{n} d_i}{t_n - t_0} \tag{1}$$

with

$$d_i = \sqrt{\sum_{j=1}^{k-1} (mp_j^{Trj_1} - mp_j^{Trj_2})^2} \tag{2}$$

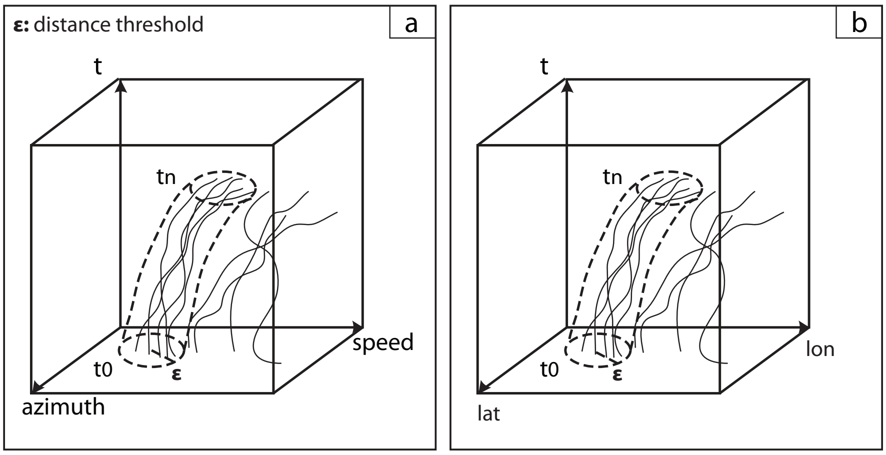
For the sake of simplicity, Figure 1.c illustrates computing the similarity between two trajectories in a three-dimensional MP space. As shown in the figure, the MP space is generated from a set of two arbitrary movement parameters (i.e. mp_1, mp_2) over time. The distance between two trajectories at timestamp t_i is d_i and is computed as the Euclidean distance between the two points in the k-dimensional MP space (Fig. 1.c and Equation 2). Where k is the number of movement parameters that are considered, with the time dimension added (i.e. movement parameters account for k-1 dimensions, plus time).

The proposed method applies the Euclidean distance, since the complexity of computing this measure is linear (i.e. $O(n)$). The measure is easy to implement and requires no control parameter. Moreover, as shown in Ding et al. (2008), the Euclidean distance can compete with more complex measures such as edit distance and LCSS in very large datasets. However, this measure is sensitive to noise and outliers. Therefore, a pre-processing step is recommended to alleviate this problem.

It is necessary to remark that prior to the similarity computation, movement parameter profiles have been normalised to the scale of one (i.e. $mp_j \in [0-1]$). Therefore, the total dissimilarity value, $D(Trj_1, Trj_2)$, is between 0 - 1. That is, if the total distance equates to one, the subject trajectories are at the maximum dissimilarity or least similarity. In contrast, dissimilarity values less than a small distance threshold (i.e. $0 < D \leq \epsilon$) indicate that the trajectories to a large extent resemble each other. The distance threshold (ϵ) is required in a majority of similarity assessment techniques, is typically application-specific and may vary depending on the purpose of the queries.

2.4 Movement Pattern Detection based on Similarity

Most movement patterns emerge from similarity in one or more movement parameters (Laube, 2005). For example, *concurrence* and *coincidence* movement patterns appear from trajectories of objects that exhibit similar movement characteristics over time (like the ones inside the 'tube' in Fig. 2). Here we suggest using the proposed similarity measure in the detection of such patterns in the MP space. *Concurrence* is defined over a period $[t_0, t_n]$ in the attribute space (e.g. (speed, azimuth, t) in Fig 2.a), whereas*coincidence* is defined over a period $[t_0, t_n]$ in the space-time cube (i.e. (lat , lon, t) in Fig 2.b). The distance between identical trajectories equals zero.



a) Concurrence is defined in attribute space (e.g. speed, azimuth, t)

b) Coincidence is defined in space-time cube (i.e. lat , lon, t)

Figure 2: Concurrence and coincidence patterns: trajectories that stay within the threshold region (tube) for a given time period $[t_0, t_n]$ are identified as patterns.

3 Case Study

The proposed methodology can be exploited for clustering trajectory data, and the discovery of spatio-temporal movement patterns, whenever the interest is to detect common movement characteristics of objects over time. In this study, we evaluate the proposed method on 100 years of historical trajectories of North Atlantic hurricanes[1] that occurred between 1907 and 2007.

3.1 Objective

This study uses hurricanes as a test bed for an exemplary proof of concept that the developed method is capable of identifying patterns of similar trajectories such as concurrence and coincidence movement patterns. Hence, this case study intends to seek for such similarity patterns in

[1]from NOAA's Coastal Services Centre (http://csc-s-maps-q.csc.noaa.gov/hurricanes/)

the movement characteristics of hurricanes, specifically, around the time of landfall. According to the meteorological literature, the most critical moment of the movement of a hurricane is at the time of recurvature (i.e. change to a more northerly direction). Moreover, the destruction caused by hurricanes happens at the time and location of the landfall (Elsner and Kara, 1999). Therefore, it is essential to gain knowledge about the behaviour of hurricanes around these two points in their evolution.

3.2 Pre-processing the Hurricane Dataset

For this study, 397 trajectories of North Atlantic hurricanes that made landfall were considered. From each hurricane trajectory a subtrajectory starting from four days before the time of landfall to 1 day after landfall was extracted. Out of the 397 only 167 hurricanes were long enough and coincided with the selected time window. 167 hurricane subtrajectories of the duration of 5 days were thus obtained. The reason for the selection of such data was that we wanted to investigate the hurricane movement patterns from around the recurvature point to shortly after the landfall. Figure 3 shows the obtained subtrajectories in dark blue, which were then used for the similarity analysis experiments. The original hurricane tracks obtained from NOAA (shown in light blue) contain little noise and are regularly sampled (i.e. at a 6 hours interval). Therefore, the pre-processing stage did not involve resampling or smoothing.

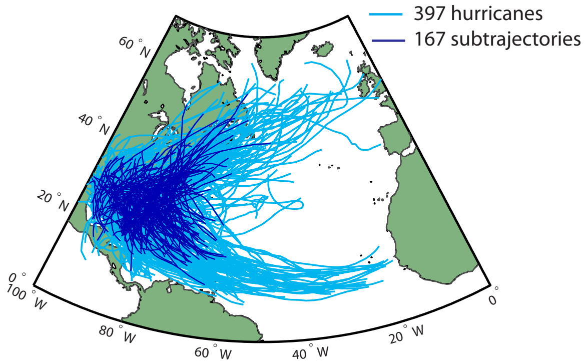


Figure 3: 397 hurricanes with landfall between 1907 and 2007; and the selected 167 hurricane subtrajectories of 5 days duration.

3.3 Similarity Assessment and Movement Pattern Detection

At this stage, for each hurricane subtrajectory the movement parameter space was generated from the speed, acceleration, azimuth, and turning angle profiles. Four (167×167) matrices of the pair-wise distances between trajectories were then computed using the proposed similarity function with different settings of the movement parameters:

1. Latitude – longitude – time: to find hurricanes that followed similar movement path (geometric shape).

2. Speed – azimuth – time: to find hurricanes that moved with similar direction and speed.

3. Speed – turning angle – time: to find hurricanes that generated similar curvature at similar speed.

4. Speed – acceleration – turning angle – time: to find hurricanes that generated similar curvature at similar speed and acceleration.

These distances were then used for the discovery of *concurrence* and *coincidence patterns*. For the purpose of pattern discovery, one arbitrarily selected sample hurricane subtrajectory was considered as a 'reference (or query) pattern'. Then, the hurricanes which attained a distance less than a small threshold to the 'reference pattern' subtrajectory were extracted.

The *coincidence* patterns were discovered on the 3D space-time cube (Fig. 2.b), computed from the geographic (lat/lon) coordinates of hurricanes over time (i.e. setting (1)), whereas the following settings were investigated for the extraction of the *concurrence* patterns:

3. Concurrence of speed and azimuth over time (Fig. 7)

4. Concurrence of speed and turning angle over time (Fig. 8)

5. Concurrence of speed, acceleration, and azimuth over time (Fig. 9)

Figures 4 and 5, respectively, illustrate the *coincident* subtrajectories (in light blue) that were extracted for two different reference patterns (shown in dark blue). For this case, after running a set of experiments the optimum threshold was set at 0.07. Figure 5 shows the effect of changing the threshold: By reducing the distance threshold from 0.07 to 0.06 four trajectories (the ones that are labelled with $*$) do not match to the reference pattern. This confirms the visual impression that their shape is less similar to the reference pattern.

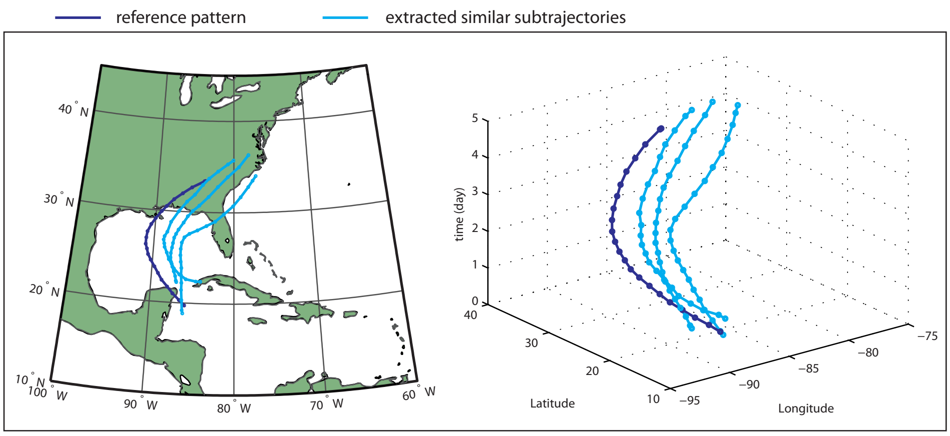


Figure 4: Extracted *coincidences* for a first reference pattern.

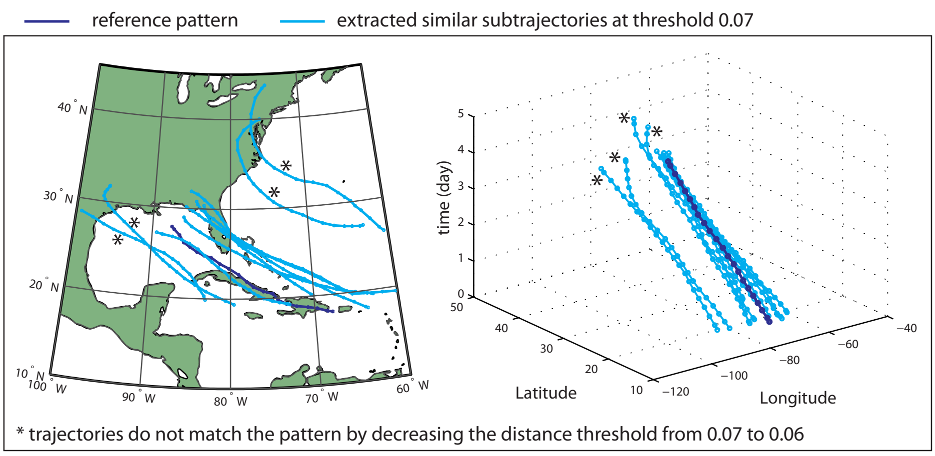


Figure 5: Extracted *coincidences* for a second reference pattern.

Figure 6 illustrates in magenta 10 % of the subtrajectories that did not match the reference pattern of Figure 5 within the selected threshold. This small subset of 10 % of all 'dissimilar' subtrajectories was chosen arbitrarily and simply to avoid over-crowding of the display. As

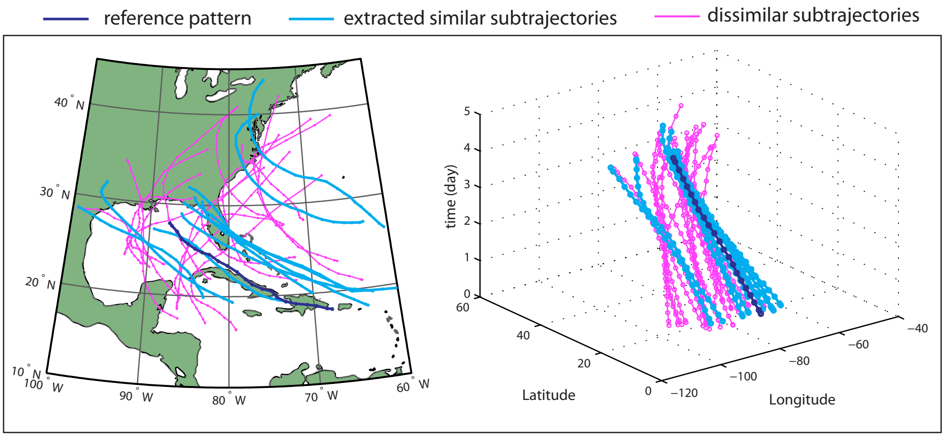


Figure 6: Extracted *coincidences* for the reference pattern of Figure 5.

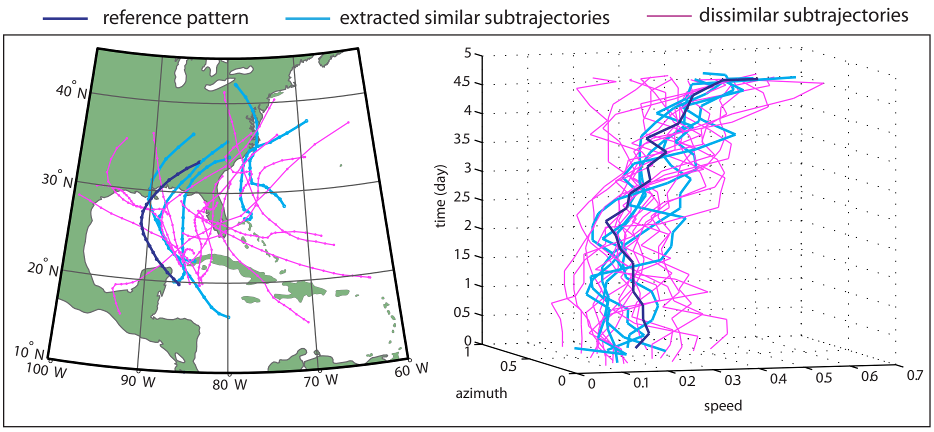


Figure 7: Extracted *concurrences* of 'speed – azimuth – time' for a reference pattern.

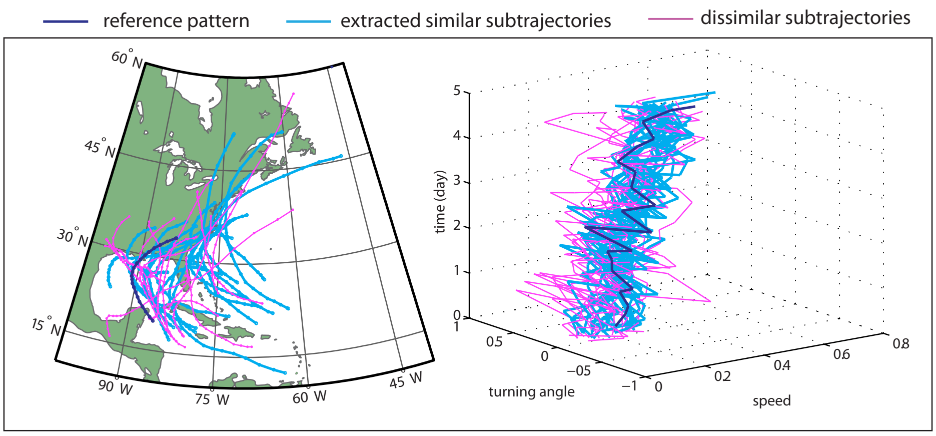


Figure 8: Extracted *concurrences* of 'speed – turning angle – time' for a reference pattern.

it can be observed, the method could successfully distinguish and excludes the subtrajectories with a geometry that is different from the sample pattern. Hence, the results suggest that the proposed similarity assessment method is useful for the detection of *coincidence* patterns. Also, the method facilitates discovering different types of *concurrence* patterns only by changing the setting of the MP space (Figures 7-9). The subtrajectories shown in magenta in the figures are an arbitrary 10 % of the subtrajectories that did not match to reference patterns. Figure 9 does not visualise the corresponding movement parameter space, since the space was four-dimensional.

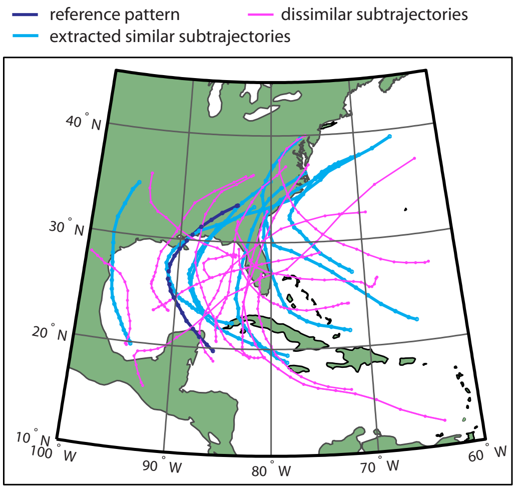


Figure 9: Extracted *concurrences* of 'speed – acceleration –turning angle – time' for a reference pattern.

According to the meteorological literature (Elsner and Kara, 1999), hurricanes that originate from proximate spatial latitudes exhibit similar movement characteristics. As a consequence, meteorologists distinguish two classes of hurricanes: 'low-latitude hurricanes' and the 'high-latitude hurricanes' with respect to the 20° N latitude. The same observation has been obtained from the hurricanes of the same season (Elsner and Kara, 1999). That is, the hurricanes of the early season (April, May, June, and July) to some extent have a dynamic behaviour that sets them apart from the late season hurricanes (August, September, October, and November).

In order to assess whether the obtained patterns confirm the meteorological hypotheses, we computed counts for the extracted hurricane subtrajectories and related these to the time of formation and the latitude of the origin of the hurricanes. The two reference patterns used in this study belonged to two arbitrary late season hurricanes (i.e. occurred in September) with origin locations below the 20° N. Table 1 summarises the counts for the extracted subtrajectories. The outcomes suggest that the extracted hurricanes of similar movement characteristics to a great extent also share the same attributes in terms of their formation time and their locations of origin. This further demonstrates the utility of the proposed technique in hurricane research.

Table 1: Counts of the extracted patterns w.r.t. the time of formation and latitude of origin of the corresponding hurricanes.

	No. extracted similar subtrajectories	**late season**	**latitude $\leq$ 20° N**
(latitude, longitude, time) shown in Figure 4	3	2	2
(latitude, longitude, time) shown in Figure 5 where threshold is 0.06	5	4	5
(speed, azimuth, time) shown in Figure 7	5	3	4
(speed, turning angle, time) shown in Figure 8	19	18	11
(speed, acceleration, turning angle, time) shown in Figure 9	8	7	5

4 Conclusions and Outlook

This paper proposed a simple, yet effective method for assessing the spatio-temporal similarity of the movement of dynamic objects and processes. We evaluated the applicability of the proposed technique through a set of experiments using trajectories of North Atlantic hurricanes. The results suggest that the proposed method can successfully reproduce the existing meteorological knowledge about the movements of hurricanes by the extraction of different movement patterns such as *coincidence* and *concurrence*. The strategy for future work is twofold: First, instead of using arbitrary query patterns, we will use the suggested similarity measures for a systematic search for coincidence and concurrence patterns (i.e. extraction of frequent movement patterns in a large dataset). Second, the influence of the distance threshold will be investigated in a systematic sensitivity study.

References

Buchin, K., Buchin, M., Gudmundsson, J., Löffler, M., and Luo, J. (2008). Detecting commuting patterns by clustering subtrajectories. In *Proceedings of the 19th International Symposium on Algorithms and Computation*, number 642, pages 644–655, Gold Coast, Australia. Springer.

Buchin, K., Buchin, M., van Kreveld, M., and Luo, J. (2009). Finding long and similar parts of trajectories. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '09*, pages 296–305, New York, New York, USA. ACM Press.

Chen, L., Özsu, M. T., and Oria, V. (2005). Robust and fast similarity search for moving object trajectories. In *SIGMOD '05: Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, pages 491–502, New York, NY, USA. ACM.

Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., and Keogh, E. (2008). Querying and mining of time series data: experimental comparison of representations and distance measures. *Proc. VLDB Endow.*, 1(2):1542–1552.

Dodge, S., Weibel, R., and Forootan, E. (2009). Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. *Computers, Environment and Urban Systems*, 33(6):419–434.

Dodge, S., Weibel, R., and Lautenschütz, A.-K. (2008). Towards a taxonomy of movement patterns. *Information Visualization*, 7(3-4):240–252.

Elsner, J. and Kara, A. (1999). *Hurricanes of the North Atlantic: Climate and society.* Oxford University Press, New York Oxford.

Giannotti, F. and Pedreschi, D. (2008). *Mobility, Data Mining and Privacy.* Springer-Verlag, Berlin Heidelberg.

Gudmundsson, J. and van Kreveld, M. (2006). Computing Longest Duration Flocks in Trajectory Data. In *GIS '06: Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems*, pages 35–42, New York, NY, USA. ACM.

Laube, P. (2005). *Analysing Point Motion, Spatio-Temporal Data Mining of Geospatial Lifelines.* PhD thesis, Universitz of Zurich.

Laube, P., Dennis, T., Forer, P., and Walker, M. (2007). Movement beyond the snapshot Dynamic analysis of geospatial lifelines. *Computers, Environment and Urban Systems*, 31(5):481–501.

Miller, H. J. and Han, J. (2009). *Geographic Data Mining and Knowledge Discovery*. Taylor & Francis Group, second edition.

Pelekis, N., Kopanakis, I., Marketos, G., Ntoutsi, I., Andrienko, G., and Theodoridis, Y. (2007). Similarity Search in Trajectory Databases. In *TIME'07: Proceedings of the 14th International Symposium on Temporal Representation and Reasoning*, pages 129–140, Washington, DC, USA. IEEE Computer Society.

Trajcevski, G., Ding, H., Scheuermann, P., Tamassia, R., and Vaccaro, D. (2007). Dynamics-aware similarity of moving objects trajectories. In *GIS '07: Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*, pages 1–8, New York, NY, USA. ACM.

Vlachos, M., Gunopulos, D., and Kollios, G. (2002). Discovering similar multidimensional trajectories. In *ICDE '02: Proceedings 18th International Conference on Data Engineering*, pages 673–684. IEEE Computer Societ.

Yanagisawa, Y., Akahani, J.-i., and Satoh, T. (2003). Shape-Based Similarity Query for Trajectory of Mobile Objects. In *MDM '03: Proceedings of the 4th International Conference on Mobile Data Management*, pages 63–77, London, UK. Springer-Verlag.

Zhang, Z., Huang, K., and Tan, T. (2006). Comparison of Similarity Measures for Trajectory Clustering in Outdoor Surveillance Scenes. In *ICPR'06: 18th International Conference on Pattern Recognition*, volume 3, pages 1135–1138, Los Alamitos, CA, USA. IEEE Computer Society.

Biography

Somayeh Dodge is a Ph.D. student and research assistant at the Department of Geography, University of Zurich. Her main research interests are mobility data mining and trajectory similarity analysis. She has been recently awarded a grant for a postdoctoral research on 'context-dependent similarity analysis of movement' at the University of Zurich.

Robert Weibel is a professor of GIScience at the Department of Geography, University of Zurich. He serves as chair of the European research coordination action COST IC0903 Knowledge Discovery from Moving Objects (MOVE), as well as on various journal editorial boards and program committees of conferences in GIScience.

Patrick is a lecturer with the Department of Geography, University of Zurich. Patrick's main research interests are movement analysis, spatio-temporal data mining, and most recently decentralised spatial computing for geosensor networks.